データ分散保存システム

石田祐子（ジャパンデータコム㈱）、井上徹（広島修道大学）、
佐久間隆（㈱コムプラス）、下田泰戈（㈱sigic）、金平恵実（(有)ソフトフィル）
宮前俊一、竹久達也、森本健嗣（㈱ニッシン）

パソコン程度の設備を使用して、暗号鍵及びデータの安全な管理を提供することによって、重要度の高いデータを保存し、 且つ暗号化によるプライバシーの保護に寄与する。 本システムでは保管すべき大量のデータはリードソロモン符号と暗号化により分散保管し、 いくつかのパソコンに保存する。その暗号鍵についてはVSS (Verifiable Secret Sharing)を用いた秘密分散で保管する。 それにより地震や火災により複数のパソコンが破壊されても確実にデータが再生できるシステムを構築した。

１．はじめに

日本政府は 2003 年までに行政の効率化や国民負担の軽減を目標に行政手続きを電子化する構想を進めている。電子政府で扱う行政のサービスデータのみならず、市井では大学における入学試験問題やデータ、企業における契約書や人事、経理情報、病院における診察や病歴ファイルなど秘密に保管すべきデータは随所に多数存在する。これらを単に暗号化しただけでは暗号化鍵の紛失や暗号データが災害などで破壊されたときデータが再現できないなどの欠陥がある。

高度で堅牢な設備を有することなく比較的安価なパソコン程度の設備を用い、暗号鍵及びデータの安全管理を提供することにより、重要度の高いデータ保管及び暗号化によるプライバシー保護に寄与する。

保管する重要かつ大量データはリードソロモン符号と暗号化により分散保管し、複数個のパソコンに保管する。なお、暗号鍵に就いてはＶＳＳ(Verifiable Secret Sharing)を用いた秘密分散で保管する。これらを実現することで大地震や火災などの災害により複数のパソコンが破滅的ダメージを受けても安全、確実にデータ再生ができるシステムを構築した。

２．研究開発の目標と内容

２．１．技術開発の目標

本システムではデータを暗号化したのち複数個（ｋ個）のブロックに分割しそれをリードソロモン符号化して n 個（n>k）のブロックに誤り訂正符号化したのち安全に保管サーバによって保管させる。このうち n-k までのサーバに障害が起きてデータが取り出せなくなってもリードソロモン復号をおこなって、もとのデータを正しく復元できる。又、暗号化に用いたデータ鍵はＶＳＳ秘密分散を用いて複数個のシェア鍵に変換され分散保管される。そのうち閾値以上のシェア鍵が収集できれば鍵が正しく再生される。このようにデータは暗号化されたあと誤り訂正符号化されて

いるので災害などで失われたデータがあっても確実にもとのデータを復元できる。又、暗号化/復号の鍵はＶＳＳ秘密分散法で分散保管されているのでそのうち閾値以上のシェア鍵を収集すればもとの鍵が正しく再生され不慮の事故に対する安全対策を万全のものとすることができる。

データ保管の安全性は、多くの箇所に分散保管することによってより確実になる。

このようなデータの保管方法としては、秘密分散（Secret Sharing）の手法が存在する。（k,N）閾値秘密分散法では、N 箇所に保管されたデータの内ｋ個のデータからもとのデータが復元される。しかし、この方法では、一つの保管場所にデータ全体と同量のデータを保管するので、N 個所にデータを保管するには元のデータの N 倍のデータを保管することになり、保管すべきデータ量が膨大な場合には不経済である。自然災害等に対するデータの安全な保管のほかに考慮すべきことは、保管データの秘密の保護である。

（k,N）閾値秘密分散法では、ｋ個の保管データからのみ元のデータが復元される。個々の保管データまたは（ｋ－１）個以下の保管データからは元のデータは復元できない。このように、（k,N）閾値秘密分散法では、データの復元性とデータの秘匿性の両方を実現している。本システムでは、データ復元のためのデータ分散保管機能と秘密保持機能を分離する。データ分散保管は、リードソロモン符号のデータの冗長性を利用して行い、秘密保持は暗号化によって実現する。さらに、暗号化に用いた暗号鍵の保管には、（k,N）閾値秘密分散法を使用する。本システムの方法では、N 箇所に分散保管されたデータの内ｋ箇所のデータから元のデータを復元できるためには、元のデータの１/k の量のデータと暗号化に必要な暗号鍵を N 箇所に分散保管すればよく、システム全体に保管されるデータ量は（N/ｋ）×（データ量）＋N×（暗号鍵の長さ）となる。通常、データ量は、暗号鍵の長さよりもはるかに大きいから、ほぼデータ量の（N/ｋ）倍の保管量で安全なデータの保管が行える。

２．２．技術開発の内容

本システムは機能として(1)管理機能、(2)保管機能、(3)データ生成復元機能からなる。

(1)管理機能：本機能は以下の機能を提供するものである。

管理初期化機能：コンソールから入力されたユーザＩＤ、パスワードと作成されたフロッピーディスク内のユーザＩＤ、パスワードとのチェックを行い、本機能を使用することが出来る者か否かを確認する。

登録削除機能：ユーザ登録・削除機能、サーバ登録・削除機能、鍵リスト一覧表示機能から構成される。

ユーザ登録・削除機能：本システムに参加する保管サーバ、管理サーバ、ユーザ ＰＣ はユーザＩＤ、パスワード、氏名等をユーザ登録情報として、ユーザ登録ファイル及びフロッピーディスクに出力する。削除が選ばれた場合には、ユーザ登録ファイルの該当レコードを削除する。

サーバ登録・削除機能：ユーザ登録・削除機能と同じくサーバの登録削除を行う。

鍵リスト一覧表示が選ばれた場合には、ＲＳＡ暗号鍵ファイルより公開鍵、ユーザＩＤ等の内容を表示する。

公開鍵提供機能：保管機能からの公開鍵情報要求により、ＲＳＡ暗号鍵ファイルより公開鍵情報を求め、保管機能に渡す。

登録サーバ提供機能：データ生成復元機能からのサーバ登録情報要求により、サーバ登録ファイルよりサーバ登録情報を求め、データ生成復元機能に渡す。
管理通信機能：データ生成復元機能からのサーバ登録情報要求により、サーバ登録情報を渡す。保管機能からの公開鍵情報要求により、公開鍵情報を渡す。
(2)保管機能：本機能は以下の機能を提供するものである。
保管初期化機能：コンソールから入力されたユーザＩＤ、パスワードと作成されたフロッピーディスク内のユーザＩＤ、パスワードとのチェックを行い、本機能を使用することが出来る者か否かを確認する。
データ保管返送機能：データ生成復元機能からのデータ保管要求により、分散データファイルの受け入れ状況をチェックし、結果をデータ保管可否通知としてデータ生成復元機能に渡す。
次にデータ生成復元機能からデータ情報を受取りデータ部分のみを抽出し、分散データファイルに出力し、結果をデータ受信通知としてデータ生成復元機能に渡す。データ生成復元機能からのデータ返送要求により、分散データファイルの保管状況をチェックし、結果をデータ返送可否通知としてデータ生成復元機能に渡す。
データ生成復元機能からのデータ送信要求により、分散データファイルより該当するレコードを抽出・編集し、データ情報としてデータ生成復元機能に渡す。
シェア鍵保管返送機能：データ生成復元機能からのシェア鍵保管要求により、シェア鍵ファイルの受け入れ状況をチェックし、結果をシェア鍵保管可否通知としてデータ生成復元機能に渡す。次にデータ生成復元機能からシェア鍵情報を受取りシェア鍵部分のみを抽出し、シェア鍵ファイルに出力し、結果をシェア鍵受信通知としてデータ生成復元機能に渡す。データ生成復元機能からのシェア鍵返送要求により、ユーザＩＤ、署名、署名データ等を受取る。このユーザＩＤを持つ公開鍵を管理機能に問合せ、受取った公開鍵情報で署名検証を行う。検証結果をシェア鍵返送可否通知としてデータ生成復元機能に渡す。
データ生成復元機能からのシェア鍵送信要求により、検証結果が正しければ、シェア鍵ファイルより該当するレコードを抽出・編集し、シェア鍵情報としてデータ生成復元機能に渡す。
保管通信機能：データ生成復元機能からの稼動状況問合せにより、当保管サーバ機の稼動状況通知を送信する。
データ保管時には、データ生成復元機能からのデータ保管要求により、データ保管可否通知を送信する。データ情報を受信する。データ受信通知を送信する。
データ復元時には、データ生成復元機能からのデータ返送要求により、データ返送可否通知を送信する。データ送信要求を受信する。データ情報を送信する。
シェア鍵保管時には、データ生成復元機能からのシェア鍵保管要求により、シェア鍵保管可否通知を送信し、シェア鍵情報を受信する。シェア鍵受信通知を送信する。
シェア鍵復元時には、データ生成復元機能からのシェア鍵返送要求により、シェア鍵返送可否通知を送信し、シェア鍵送信要求を受信する。シェア鍵情報を送信する。

(3)データ生成復元機能：本機能は以下の機能

を提供するものである。

データ生成復元初期化機能：コンソールから入力されたユーザＩＤ、パスワードとで作成されたフロッピーディスク内のユーザＩＤ、パスワードとのチェックを行い、本機能を使用することが出来る者か否かを確認する。

管理機能へのサーバ登録情報要求により、サーバ登録情報を受取る。

システム起動時には、保管機能への稼動状況問合せにより、稼動状況通知を受取り、既に受取ってあるサーバ登録情報に稼動状況を反映し、保管先を確定する。

シェア鍵復元機能：コンソールから入力された鍵ＩＤをキーとして、フロッピーディスクまたは保管履歴ファイルから該当する鍵保管先を求め、確定した保管先とを比較する。その結果よりシェア鍵の取出が可能か否かを判断し処理の続行、中止を決める。

処理続行と判断された場合には、保管機能へのシェア鍵返送要求により、シェア鍵返送可否通知を受取る。保管機能へのシェア鍵送信要求により、シェア鍵情報を受取る。

受取ったシェア鍵情報に付加されているチェック情報をチェックし改竄の有無を確認し、鍵復元演算により復元鍵を算出し、フロッピーディスクに出力する。

データ復元機能：コンソールから入力されたファイル名をキーとして、保管履歴ファイルから該当するファイル保管先を求め、システム起動時に確定した保管先とを比較する。その結果よりデータ情報の取出が可能か否かを判断し処理の続行、中止を決める。

処理続行と判断された場合には、保管機能へのデータ返送要求により、データ返送可否通知を受取る。保管機能へのデータ送信要求により、データ情報を受取る。

受取ったデータ情報を入力とし、リードソロモン復号にて誤り訂正復号データを作成し、暗号アルゴリズムにより復号した後、復元データファイルに出力する。

データ・鍵保管機能：乱数を生成し、ブロック暗号鍵として切り出し鍵、分散鍵、チェック情報を求める。コンソールから入力された保管対象データファイルのファイル名にて保管機能に対し、シェア鍵保管要求を行い、保管機能からシェア鍵保管可否通知を受取った後、シェア鍵情報を保管機能に渡し、保管を委ねる。

保管機能よりシェア鍵受信通知を受取ったならば、シェア鍵情報よりシェア鍵保管履歴情報を抽出し、保管履歴ファイル及びフロッピーディスクに出力する。

次に、コンソールから入力されたファイル名にて保管対象データファイルを読込み、既に求めてある切り出し鍵を用い、暗号アルゴリズムにて暗号化データを作成する。暗号化データを入力とし符号化データフォーマットに整形する。符号化データフォーマットに整形された暗号化データを入力としリードソロモン符号化を行いチェックデータを付加し、符号化データを作成する。

保管機能に対し、データ保管要求を行い、保管機能からデータ保管可否通知を受取った後、データ情報を保管機能に渡し、保管を委ねる。

保管機能よりデータ受信通知を受取ったならば、データ情報よりデータ保管履歴情報を抽出し、保管履歴ファイルに出力する。

データ生成復元通信機能：管理機能へのサーバ登録情報要求にて、サーバ登録情報を受信する。システム起動時には、保管機能への稼動状況問合せにて、稼動状況通知を受信する。シェア鍵復元時には、保管機能へのシェア鍵

返送要求により、シェア鍵返送可否通知を受信する。保管機能へのシェア鍵送信要求により、シェア鍵情報を受信する。
データ復元時には、保管機能へのデータ返送要求により、データ返送可否通知を受信する。保管機能へのデータ送信要求により、データ情報を受信する。

データ・鍵保管時には、保管機能へのシェア鍵保管要求により、シェア鍵保管可否通知を受信する。保管機能へシェア鍵情報を送信し、シェア鍵受信通知を受信する。

保管機能へのデータ保管要求により、データ保管可否通知を受信する。保管機能へデータ情報を送信し、データ受信通知を受信する。

シェア鍵の生成/復元には、ＶＳＳ鍵分散保管機能を用いる。これはデータを暗号化する鍵を秘密分散法ＶＳＳ（Verifiable Secret Sharing）を用いて分散保管する。具体的には鍵をm個のシェア鍵に変換し、保管サーバに保管させる。このうち閾値 a 個（m>a）のシェア鍵が収集できれば元の暗号化に使った鍵が再現できる。データの暗号化機能は、データ暗号復号機能を用いる。

本機能は暗号化モードでは 64 ビット単位の暗号化を行い、復号モードでは64 ビット単位の復号を行う。保管すべきデータを DES または MISTY により暗号化/復号する機能である。データの誤り訂正符号化/復号には誤り訂正符号化/復号機能を用いる。
誤り訂正符号であるリードソロモン符号を使ってｋ個にブロックわけされたデータを誤り訂正符号化してｎ個のブロックに符号化する。ここで n>k である。

本機能は保管すべきデータを整形し暗号化したのち、その後誤り訂正符号化する機能を提供する。又、保管データを読み出し符号化データを再構成しこれにリードソロモン符号の復号を行う。この時データに誤りが検出されると訂正能力内であれば訂正する。基本機能として RSA の鍵生成には基本鍵生成機能を用いる。

本機能はユーザの公開鍵のペア鍵を生成する。初期化機能から乱数を生成し、素数判定を行って素数ｐ、ｑを生成する。暗号鍵ｅ，復号鍵ｄを生成する。署名生成には署名生成検証機能を用いる。

本機能はセション開始時相互認証のために、保管データ生成 PC より確認テキストに署名をして保管 PC へ送信する。保管 PC において確認し正常であれば、保管 PC にて確認テキストに署名し保管データ生成 PC へ送信する。互いに署名認証と検証を行う。セッション開始時は相互認証として保管データ生成ＰＣより確認テキストに送信する。この時点でセッションが確立する。この演算には多倍長精度演算機能を用いる。

本機能はＲＳＡなどの演算は 1024 ビット演算を行うため通常の演算ソフトでは実行できない。1024 ビット演算まで可能にする。
通信機能は各種の初期化、パソコン間の送信受信、システムの障害検知、各プロセスの終了などを行う。
初期化機能は各種データベースの初期化を行う。
送信受信機能は送信受信機能を提供する。データ保管モードではＶＳＳ機能の出力パラメータを各保管パソコンへ送信する機能を提供する。再構成局面では相互認証を行い通信相手を確認する。
ＦＤ機能は保存情報をフロッピーディスクに書込んだり、フロッピーディスクから読込んだりする。

３．本年度の活動状況

通信プロトコル技術としてはセッション層、プレゼンテーション層、アプリケーション層に独自のプロトコルを開発し、所定の動作を確認した。

開発成果物の稼動環境：

開発するソフトウェアは、以下の環境で稼動させた。

（１）管理サーバ

・ハード：Intel PentiumⅢ、500MHZ、64MB～(Memory)、10GB～(HD)

・搭載ＯＳ：Linux（Red Hat　Ver7,2J）

（２）ユーザＰＣ（データ生成復元 PC）

・ハード：Intel PentiumⅢ、800MHZ、128MB～(Memory)、20GB～(HD)

・搭載ＯＳ： Windows2000

（３）保管サーバ

・ハード：Intel PentiumⅢ、800MHZ、128MB～(Memory)、20GB～(HD)

・搭載ＯＳ：Linux（Red Hat　Ver7.2J）

分散パラメータ n および k:

ユーザ PC（データ生成復元 PC）はソフトウェアとしては制限なし。取りうるパラメータは$n \leqq 256$、$1 \leqq k < n$のあらゆる組み合わせを取りうるが、もっとも推奨する組み合わせは n=5,k=3 である。理由は今日 Linux であれば 5,6 台なら数十万円で構成できる。又、ハードディスクが 2 台クラッシュしてもシステムに影響しない。同じことをミラーサーバーで行うと冗長度が５０％になるし、ハードディスクが２台クラッシュするとシステムダウンする可能性がある。

本システムは４０％の冗長度で済む。さらに 3 台のディスクが同時にクラッシュする可能性は少ない。2 台までで障害に気づいて直せる確率が高い。

クライアント C-job によるタスク管理：

図 2 に示すようにデータ生成復元機能を実現する際、ソフトウエアからみて６つのフェーズに分割し、それぞれユーザインターフェイス(CUserInterface)クラス、暗号化(CEncryption)クラス、復号(CDecryption)クラス、分散(CScatter)クラス、集合(CGather)クラス、通信(CCommunication)クラスとし、各クラス（フェーズ）を呼び出すためのデータ保存領域として作業(CJob)クラスを定義する。

また、データ生成復元機能の各機能を実行する為に、ユーザからの指示により起動される一連の処理を実行するセッション(CSession)クラスを定義する。

データ生成復元機能がデータ保管する際には、ユーザ操作を抽象化した CUserInterface クラスが CJob インスタンスを作成し、CJob に対してデータ保管を指示した後、その他必要な情報をセットし、CSession クラスへ引き渡す。受け取った CSession クラスは、受け付けた CJob クラスにデータ保管と記載があるので、その CJob インスタンスに対して、暗号化を指示し、暗号化に必要な情報をセットした後、CEncryption へ引き渡す。CEncryption の処理が正常に終了したことを CJob クラス経由で、CSession クラスが知ると、次に CJob クラスに対して分散化を指示し、必要な情報をセットした後、CScatter へ引き渡す。CScatter の処理が正常に終了したことを CJob クラス経由で、CSession クラスが知ると、次に CJob クラスに対して、データ送信を指示し、必要な情報をセットした後、CCommunication クラスへ引き渡す。CCommunication の処理が正常に終了したことを CJob 経由で、CSession クラスが知ると、

CJob に対して、終了状況をセットし、CUserInterfaceに終了したことを知らせ、データ保管機能が実行終了となる。

他の機能も、CUserInterface からの CJob 経由で指示を CSession へ送り、CSession が要求の指示内容に沿ったフェーズを順に呼び出す。こうすることで、他の機能を簡単に付加することが出来るような設計にしている。

サーバはクライアントからの接続を受け付け、要求に対して応答する。クライアントは、サーバへ接続を行い、要求を送り、要求に対する返答を受け取るという動作を行う。各プロセスは、Socket インターフェイスといわれるインターフェイスに準じた形で作成することにより、他機種、他環境でもほぼ同じプログラムで動作させることが出来るように移植性を考慮し作られている。

通信プロトコルは RFC STD7 準拠の TCP プロトコルでソケットはストリーム型を使う。ポート番号を設定ファイルで定義する。通信データはビッグエンディアンで格納する。データ形式は S-JIS とする。応答待ち時間は default で 500ms とする。

fork による通信 job 振り分け：

fork はプロセスを生成する、Linux（UNIX）のシステム命令である。Linux はかなりはっきりしたプロセスをいくつも同時に持つことができるので、今回のシステムではサーバ機能で、外部よりランダムに要求がくるため、これに対する処理はそれぞれプロセスを発生させてそこで対応する。

４．外部発表および成果物

学会等への外部発表論文はなし。

IPA に対する本年度の納入物件：

（１）開発成果報告書　一部

〔ソフトウェア開発編〕

・外部仕様書

・試験仕様書／報告書

〔取扱説明編〕

・取扱説明書

（２）電子媒体　一式

・ソースプログラム

・ロードモジュール

・開発成果報告書

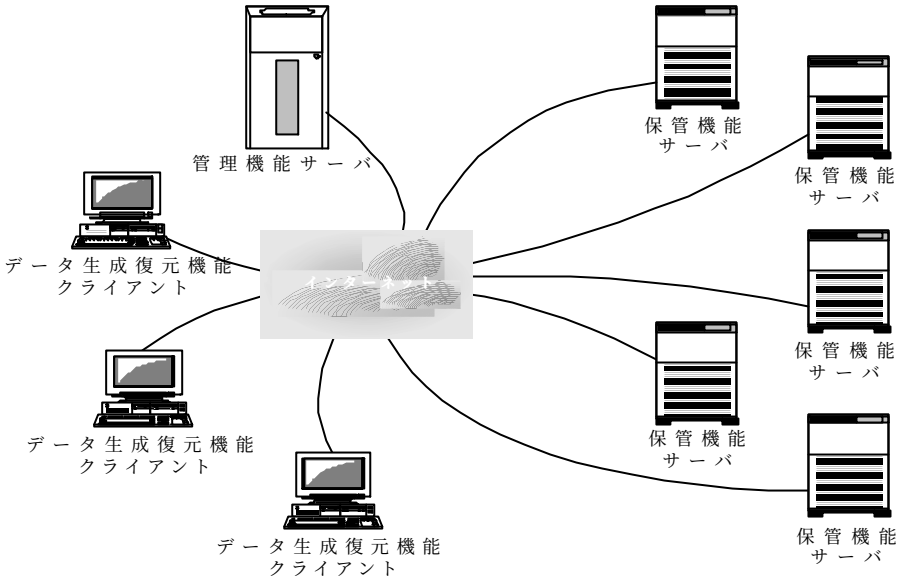

図１．システム構成図

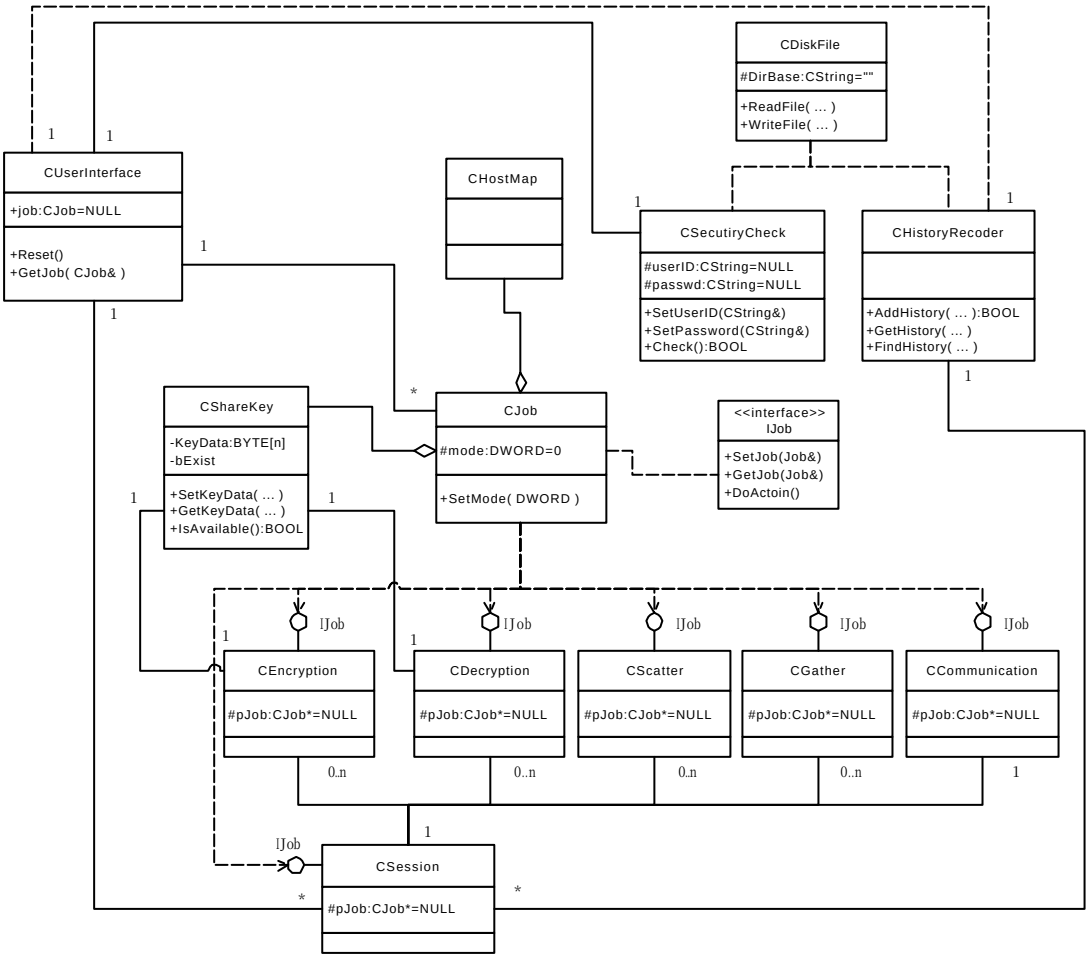

図２．C-job によるタスク管理

５．今後の課題

５．１．ST 評価

本来は開発の各段階で評価を受けフィードバックさせながら開発を進めることが必須である。提案書レベルでは基本設計のみの提案で開発に余裕がなく、限られたリソースで開発と ST 評価の仕事を並列に進めるには工数が増加して単年度では不可能であった。

５．２．プログラムも JAVA 化

プログラムの JAVA 化を行う。すなわちクライアントのプログラム部分を JAVA で記述し、管理サーバの Web ファイルにおいておく。クライアントは随時、保管してもらいたいデータが発生したらその時点で管理サーバにアクセスし、認証してもらいデータを保管してもらう。これにより常時は不要なプログラムをクライアントパソコンに蓄積せず、必要なときプログラムをダウンロードし、必要なだけのデータを保管サーバへ送信してデータを安全に保管できる。これにより移動端末を操作していても重要データ発生時は即時、データ保管できるようになる。

５．３．システム融通性の増大

保管パソコンの数を管理機能で管理しつつ設定値ｎより多くし（ｎ´＞ｎ）ダイナミックに誤り訂正能力を高めてダウンしたパソコンの数が増えてもダイナミックに障害を受けたデータを復元する能力を高めること、などをおこなう。

５．４．メモリーカード機能

フロッピーをメモリーカードとすることによりより携帯性が増し、安全性を増すことが出来る。

６．まとめ

本年度はデータの分散保存システムの方式を設計し、商用モデルシステムを構築し、基本的機能の確認とシステムの設計書の作成、暗号化/復号、誤り訂正符号化/復号の処理時間のデータを取った。

本システムはデータ、暗号復号鍵ともに最適な方法で分散保管し、災害や盗難などからシステムを守る重要且つ基本的な技術であり、クライアントプログラムの JAVA 化により更なる利便性、発展が見込まれる有望な技術である。

参考文献：

(1)Hideki IMAI(Co-authored T.INOUE et al)," Essentials of Error-Control Coding Techniques", Academic Press,Inc. 1990.

(2)井上徹、松井充、山岸篤弘、吉田英夫，長谷川広樹、三宅真、古川輝雄“実戦　誤り訂正技術”、トリケップス、１９９３年１２月。

(3)西沢台二、田崎三郎、“「先端技術の手ほどきシリーズ　ディジタル放送」第６章、ディジタル　放送と誤り訂正技術、
テレビジョン学会編、監修、オーム社。

(4) A. Shamir, "How to Share a Secret", Communications of the ACM, 22(1979), pp. 612-613.

(5)E. R. Berlek amp, Algebraic coding theory, McGraw-Hill, Inc, 1968.

(6) B. Pfitzmann: "Digital Signature Schemes", Springer-Verlag, 1996

(7) A. Salomaa: "Public-key Cryptography", Springer-Verlag, 1990

(8)Douglas R. Stinson: "CRYPTOGRAPHY:

Theory and Practice", CRC Press, 1995
(9)Bruce Schneier: "APPLIED CRYPTOGRAPHY", Second Edition, John Wiley and Sons, Inc., 1996
(10)R. L. Rivest: A. Shamir, and L. M. Adelman, "On Digital Signatures and Public Cryptosystems", Communications of the ACM, vol. 21, no. 2, Feb. 1978